1. DIFFERENTIAL PRIVACY FROM STATISTICAL QUERY ALGORITHMS

If $C$ is efficiently learnable from SQ’s, then $C$ is efficiently PAC-learnable, differential-privately

**Theorem 1.** Suppose some algorithm $A$ efficiently learns $C$ to error $\varepsilon$ from $M$ statistical queries of tolerance $\tau$. Then some algorithm $B$ efficiently PAC-learns $C$ to error $\varepsilon$ with probability $\geq 1 - \delta$ while satisfying $\alpha$-differential privacy, using

$$O\left(\frac{M}{\alpha \tau} + \frac{M}{\tau^2} \ln \frac{2M}{\delta}\right)$$ samples

**Proof.** Algorithm $B$ draws $O\left(\frac{M}{\alpha \tau} + \frac{M}{\tau^2} \ln \frac{2M}{\delta}\right)$ random samples (call them $S$)

Break $S$ into $M$ disjoint chunks $S_1, \ldots, S_M$, each of size $O\left(\frac{1}{\alpha \tau} + \frac{1}{\tau^2} \ln \frac{2M}{\delta}\right)$

Answer $i$-th statistical query $(\varphi_i, \tau)$ of $A$ using $S_i$ (taking average of $\varphi_i$ over $S_i$)

To each response, add Laplacian noise of scale $M/|S_i|\alpha$

Finally return $A$’s hypothesis $h$

**Privacy:** Each query is the average of $|S_i|$ values, each between 0 and 1

By Theorem in Notes20, each response satisfies $\alpha/M$-differential privacy

By Composition property, the collection of all $M$ responses satisfies $\alpha$-differential privacy

**Error:** By Hoeffding, with prob $\geq 1 - \delta/(2M)$,

- empirical average of $\varphi_i$ over $S_i$ (before adding noise) is within $\tau/2$ of the true expectation

With prob $\geq 1 - \delta/(2M)$, the Laplace noise has magnitude

$$O\left(\frac{1}{\alpha |S_i|} \ln \frac{2M}{\delta}\right) \leq \frac{\tau}{2}$$ since $|S_i| \geq \frac{C}{\alpha \tau} \ln \frac{2M}{\delta}$ for some large $C$

Hence with prob $\geq 1 - \delta/M$, the $i$-th response $\hat{P}_{\varphi_i}$ is within $\tau$ of the true expectation $P_{\varphi_i}$

By union bound over all $M$ queries, with prob $\geq 1 - \delta$

- all responses are within $\tau$ of their true averages, and algorithm $A$ succeeds

2. GEOMETRIC MECHANISM

When response of STAT$(c, D)$ is integer-valued, geometric mechanism may be used

**Geometric mechanism** adds noise that is a (symmetric) geometric random variable

(Symmetric) **geometric distribution** with parameter $\alpha > 1$ has pmf $f(k) = \alpha^{-|k|}(1 - \alpha)/(1 + \alpha)$

Like the Laplace distribution, symmetric geometric distribution changes by (at most) the same multiplicative factor when shifted, i.e.

$$f(k + j)/f(k) = \alpha^{-|k-j|}/\alpha^{-|k|} \leq \alpha^{|j|}$$ for any $j, k \in \mathbb{Z}$

In fact the distribution is defined so that this inequality is achieved as an equality for certain $j, k$

If symmetric geometric noise with parameter $\alpha$ is added to the output of an integer-valued function $g$

Then the mechanism satisfies $\varepsilon$-differential privacy where $e^\varepsilon = \alpha^{\Delta g}$ (exercise)

Again $\Delta g = \text{maximum change to } g$’s output when just one data point changes

By the same calculations as the Laplace mechanism

In practice a response of STAT$(c, D)$ may be required to bounded, say between 0 and $b$

Can **truncate** the response to force it to lie in the desired range, without hurting privacy (exercise)